

日 本 国 特 許 庁
JAPAN PATENT OFFICE

別紙添付の書類に記載されている事項は下記の出願書類に記載されている事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed with this Office.

出 願 年 月 日 2 0 0 3 年 3 月 6 日
Date of Application:

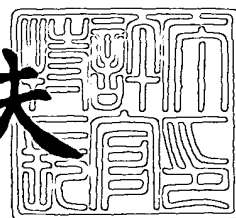
出 願 番 号 特 願 2 0 0 3 - 0 6 0 6 8 1
Application Number:
[ST. 10/C]: [J P 2 0 0 3 - 0 6 0 6 8 1]

出 願 人 財団法人名古屋産業科学研究所
Applicant(s):

2 0 0 3 年 8 月 2 8 日

特許庁長官
Commissioner,
Japan Patent Office

今 井 康 夫



出証番号 出証特 2 0 0 3 - 3 0 7 0 4 0 6

【書類名】 特許願

【整理番号】 P03-007

【提出日】 平成15年 3月 6日

【特記事項】 特許法第 3 0 条第 1 項の規定の適用を受けようとする特
許出願

【あて先】 特許庁長官殿

【国際特許分類】 G06F 9/44
G06F 17/27

【発明者】

 【住所又は居所】 愛知県名古屋市千種区鹿子町 2 - 1 4 - 2

 【氏名】 稲垣 康善

【発明者】

 【住所又は居所】 愛知県名古屋市千種区光が丘 2 - 3 - 5 シャトーひか
り 2 0 1 号

 【氏名】 松原 茂樹

【発明者】

 【住所又は居所】 岐阜県土岐市肥田町肥田 2 8 7 - 3 3 5

 【氏名】 加藤 芳秀

【発明者】

 【住所又は居所】 愛知県名古屋市港区入場 1 - 4 2 0

 【氏名】 湊 恵一

【特許出願人】

 【識別番号】 598091860

 【氏名又は名称】 財団法人 名古屋産業科学研究所

【代理人】

 【識別番号】 100110744

 【弁理士】

 【氏名又は名称】 藤川 敬知

 【電話番号】 (052)800-8122

【手数料の表示】**【予納台帳番号】** 136479**【納付金額】** 21,000円**【提出物件の目録】****【物件名】** 明細書 1**【物件名】** 図面 1**【物件名】** 要約書 1**【プルーフの要否】** 要

【書類名】 明細書

【発明の名称】 有限状態変換器作成装置、プログラム、記録媒体、作成方法、及び漸進的構文解析装置

【特許請求の範囲】

【請求項 1】 漸進的構文解析に用いる有限状態変換器を作成する装置であって、

文脈自由文法に基づく文法規則の集合を表すネットワークの集合であると共に、前記各ネットワークにおける非終端記号による遷移が他のネットワークによって定義される再帰的構造を有する再帰遷移ネットワークを作成する再帰遷移ネットワーク作成手段と、

開始記号を入力ラベルとする弧を持つ有限状態変換器を初期の有限状態変換器とし、前記有限状態変換器の弧をその入力ラベルに対応した前記再帰遷移ネットワーク中のネットワークで置き換え、さらに、その置き換えによって新たに作成された弧を、前記再帰遷移ネットワーク中の別のネットワークに置き換える操作を再帰的に繰り返す弧置き換え手段と、

文法規則の適用頻度に関する統計情報に基づいて、前記有限状態変換器における入力ラベルが非終端記号である全ての弧について各々に対応する構文本の節点の導出確率を計算し、得られた導出確率を弧の置き換え優先度とする優先度計算手段と、

を備え、

前記弧置き換え手段は、前記弧の置き換え優先度が高い弧から順に置き換え操作を適用すると共に、前記置き換え操作が繰り返し適用されることによって前記有限状態変換器が所定の大きさに達したときに前記弧の置き換え操作の適用を終了することを特徴とする有限状態変換器作成装置。

【請求項 2】 前記弧置き換え手段による前記弧の置き換え操作の適用が終了した後、非終端記号を入力ラベルに持つ弧を除去しつつ、弧の置き換え操作の適用をさらに実行する弧除去手段、を備えたことを特徴とする請求項 1 に記載の有限状態変換器作成装置。

【請求項 3】 前記節点の導出確率は、構文本における開始記号から対象の節

点までのパス上の各節点について順に文法規則が適用される確率であることを特徴とする請求項 1 又は 2 に記載の有限状態変換器作成装置。

【請求項 4】 漸進的構文解析に用いる有限状態変換器を作成するためにコンピュータを、

文脈自由文法に基づく文法規則の集合を表すネットワークの集合であると共に、前記各ネットワークにおける非終端記号による遷移が他のネットワークによって定義される再帰的構造を有する再帰遷移ネットワークを作成する再帰遷移ネットワーク作成手段、

開始記号を入力ラベルとする弧を持つ有限状態変換器を初期の有限状態変換器とし、前記有限状態変換器の弧をその入力ラベルに対応した前記再帰遷移ネットワーク中のネットワークで置き換え、さらに、その置き換えによって新たに作成された弧を、前記再帰遷移ネットワーク中の別のネットワークに置き換える操作を再帰的に繰り返す弧置き換え手段、及び

文法規則の適用頻度に関する統計情報に基づいて、前記有限状態変換器における入力ラベルが非終端記号である全ての弧について各々に対応する構文木の節点の導出確率を計算し、得られた導出確率を弧の置き換え優先度とする優先度計算手段として機能させるための有限状態変換器作成プログラムであって、

前記弧置き換え手段は、前記弧の置き換え優先度が高い弧から順に置き換え操作を適用すると共に、前記置き換え操作が繰り返し適用されることによって前記有限状態変換器が所定の大きさに達したときに前記弧の置き換え操作の適用を終了することを特徴とする有限状態変換器作成プログラム。

【請求項 5】 請求項 4 に記載の有限状態変換器作成プログラムを記録したコンピュータ読み取り可能な記録媒体。

【請求項 6】 漸進的構文解析に用いる有限状態変換器を作成する方法であって、

文脈自由文法に基づく文法規則の集合を表すネットワークの集合であると共に、前記各ネットワークにおける非終端記号による遷移が他のネットワークによって定義される再帰的構造を有する再帰遷移ネットワークを作成する再帰遷移ネットワーク作成ステップと、

開始記号を入力ラベルとする弧を持つ有限状態変換器を初期の有限状態変換器とし、前記有限状態変換器の弧をその入力ラベルに対応した前記再帰遷移ネットワーク中のネットワークで置き換え、さらに、その置き換えによって新たに作成された弧を、前記再帰遷移ネットワーク中の別のネットワークに置き換える操作を再帰的に繰り返す弧置き換えステップと、

文法規則の適用頻度に関する統計情報に基づいて、前記有限状態変換器における入力ラベルが非終端記号である全ての弧について各々に対応する構文木の節点の導出確率を計算し、得られた導出確率を弧の置き換え優先度とする優先度計算ステップと、

を備え、

前記弧置き換えステップにおいて、前記弧の置き換え優先度が高い弧から順に置き換え操作を適用すると共に、前記置き換え操作が繰り返し適用されることによって前記有限状態変換器が所定の大きさに達したときに前記弧の置き換え操作の適用を終了することを特徴とする有限状態変換器作成方法。

【請求項 7】 前記弧置き換えステップにおける前記弧の置き換え操作の適用が終了した後、非終端記号を入力ラベルに持つ弧を除去しつつ、弧の置き換え操作の適用をさらに実行する弧除去ステップ、を備えたことを特徴とする請求項 6 に記載の有限状態変換器作成方法。

【請求項 8】 前記節点の導出確率は、構文木における開始記号から対象の節点までのパス上の各節点について順に文法規則が適用される確率であることを特徴とする請求項 6 又は 7 に記載の有限状態変換器作成方法。

【請求項 9】 漸進的に構文解析を行うように構成された構文解析装置であって、

請求項 6 乃至 8 のいずれかに記載の方法によって作成された有限状態変換器と

、

その有限状態変換器へ単語を入力する度に状態遷移に伴って出力される構文木を順次接続する接続処理手段と、

を備えたことを特徴とする漸進的構文解析装置。

【発明の詳細な説明】

【0 0 0 1】

【発明の属する技術分野】

本発明は、実時間音声言語処理システム等における漸進的構文解析に用いる有限状態変換器作成装置、プログラム、記録媒体、作成方法及び漸進的構文解析装置に関する。

【0 0 0 2】

【従来の技術】

同時通訳システムなどの実時間音声言語処理システムでは、ユーザの発話に対して同時にその内容を理解し、応答する必要がある。そのような処理を実現するためには、文全体が入力されるまで待ってから解析処理を行うのではなく、発話の断片が入力されるごとに、順次、解析処理を実行するという文の漸進的な解釈が重要となる。

文の構文的関係を漸進的に理解する枠組みとして、これまでに、漸進的構文解析が研究されている。漸進的構文解析では、発話の途中段階においても、それまでに入力された文の断片に対する構文木を生成する。そのため、文全体が入力されていなくとも、その時点での構文構造を理解することができる。漸進的構文解析手法として、これまでにMatsubaraらは、漸進的チャート解析手法を提案している（非特許文献 1 参照。）。この手法では、単語が入力されるごとに、入力された単語に対して文脈自由文法の文法規則を適用する操作を繰り返して、単語に対する構文木を生成し、これを文の断片に対する構文木と結合することにより、漸進的な解析処理を実現している。しかしながら、漸進的チャート解析手法では、実時間言語処理システムにおいて要求される実時間性について十分な性能が得られないという問題があった。

そこで、発明者らは、漸進的チャート解析手法における上述した問題点に鑑みて、有限状態変換器を用いた漸進的構文解析手法を提案している（非特許文献 2 参照）。この解析手法によれば、文脈自由文法を近似変換した有限状態変換器を用いて構文解析を実行するため、高速な構文解析処理を実現可能である。

【0 0 0 3】

【非特許文献 1】

S.Matsubara, et al., "Chart-based Parsing and Transfer in Incremental Spoken Language Translation", Proceedings of NLPRS'97, pp.521-524(1997)

【非特許文献 2】

湊 他、"有限状態変換器を用いた漸進的構文解析"、平成 13 年度電気関係学会東海支部連合大会論文集、P. 279 (2001)

【0004】

【発明が解決しようとする課題】

しかしながら、上述した従来の文脈自由文法を近似変換した有限状態変換器を用いた漸進的構文解析手法では、近似変換の結果、もとの文脈自由文法では解析できる文が、有限状態変換器では解析できない場合があるという問題があった。すなわち、漸進的構文解析に用いる有限状態変換器は、文法規則を表現するネットワークにより弧を再帰的に置き換えることにより作成されるが、實際上、有限状態変換器を実現するために使用されるコンピュータの記憶領域の大きさに制限があるために文解析に十分な回数の弧の置き換えができない場合があり、このため、もとの文脈自由文法では解析可能であった文が有限状態変換器では解析不能となる場合が生じていたのである。

本発明は、かかる課題に鑑みてなされたものであり、その目的は、より多くの文について漸進的に構文解析可能な有限状態変換器の作成装置、プログラム、記録媒体、作成方法、及び漸進的構文解析装置を提供することである。

【0005】

【課題を解決するための手段】

この目的を達成するために、請求項 1 に記載の有限状態変換器作成装置は、漸進的構文解析に用いる有限状態変換器を作成する装置であって、文脈自由文法に基づく文法規則の集合を表すネットワークの集合であると共に、前記各ネットワークにおける非終端記号による遷移が他のネットワークによって定義される再帰的構造を有する再帰遷移ネットワークを作成する再帰遷移ネットワーク作成手段と、開始記号を入力ラベルとする弧を持つ有限状態変換器を初期の有限状態変換器とし、前記有限状態変換器の弧をその入力ラベルに対応した前記再帰遷移ネットワーク中のネットワークで置き換え、さらに、その置き換えによって新たに作

成された弧を、前記再帰遷移ネットワーク中の別のネットワークに置き換える操作を再帰的に繰り返す弧置き換え手段と、文法規則の適用頻度に関する統計情報に基づいて、前記有限状態変換器における入力ラベルが非終端記号である全ての弧について各々に対応する構文木の節点の導出確率を計算し、得られた導出確率を弧の置き換え優先度とする優先度計算手段と、を備え、前記弧置き換え手段は、前記弧の置き換え優先度が高い弧から順に置き換え操作を適用すると共に、前記置き換え操作が繰り返し適用されることによって前記有限状態変換器が所定の大きさに達したときに前記弧の置き換え操作の適用を終了することを特徴とする。

従って、請求項 1 に記載の有限状態変換器作成装置によれば、文法規則の適用頻度に関する統計情報に基づく弧の置き換え優先度が高い弧から順に置き換え操作が適用されるので、限られた大きさで、より多くの文を解析可能な有限状態変換器を、確実に作成することができる。

【0006】

また、請求項 2 に記載の有限状態変換器作成装置は、前記弧置き換え手段による前記弧の置き換え操作の適用が終了した後、非終端記号を入力ラベルに持つ弧を除去しつつ、弧の置き換え操作の適用をさらに実行する弧除去手段、を備えたことを特徴とする。

従って、請求項 2 に記載の有限状態変換器作成装置によれば、構文解析時に使用されない非終端記号を入力ラベルに持つ弧を除去しつつ、さらに弧の置き換えを行うので、より一層多くの文を解析可能な有限状態変換器を確実に作成することができる。

【0007】

また、請求項 3 に記載の有限状態変換器作成装置は、前記節点の導出確率が、構文木における開始記号から対象の節点までのパス上の各節点について順に文法規則が適用される確率であることを特徴とする。

従って、請求項 3 に記載の有限状態変換器作成装置によれば、構文木における開始記号から対象の節点までのパス上の各節点について順に文法規則が適用される確率を弧の置き換え優先度として用いて弧の置き換え操作を行うことによって

、より多くの文を解析可能な有限状態変換器を確実に作成することができる。

【0 0 0 8】

また、請求項 4 に記載の有限状態変換器作成プログラムは、漸進的構文解析に用いる有限状態変換器を作成するためにコンピュータを、文脈自由文法に基づく文法規則の集合を表すネットワークの集合であると共に、前記各ネットワークにおける非終端記号による遷移が他のネットワークによって定義される再帰的構造を有する再帰遷移ネットワークを作成する再帰遷移ネットワーク作成手段、開始記号を入力ラベルとする弧を持つ有限状態変換器を初期の有限状態変換器とし、前記有限状態変換器の弧をその入力ラベルに対応した前記再帰遷移ネットワーク中のネットワークで置き換え、さらに、その置き換えによって新たに作成された弧を、前記再帰遷移ネットワーク中の別のネットワークに置き換える操作を再帰的に繰り返す弧置き換え手段、及び文法規則の適用頻度に関する統計情報に基づいて、前記有限状態変換器における入力ラベルが非終端記号である全ての弧について各々に対応する構文本の節点の導出確率を計算し、得られた導出確率を弧の置き換え優先度とする優先度計算手段として機能させるための有限状態変換器作成プログラムであって、前記弧置き換え手段は、前記弧の置き換え優先度が高い弧から順に置き換え操作を適用すると共に、前記置き換え操作が繰り返し適用されることによって前記有限状態変換器が所定の大きさに達したときに前記弧の置き換え操作の適用を終了することを特徴とする。

従って、コンピュータによって、請求項 4 に記載の有限状態変換器作成プログラムを実行することにより、文法規則の適用頻度に関する統計情報に基づく弧の置き換え優先度が高い弧から順に置き換え操作が適用されるので、限られた大きさで、より多くの文を解析可能な有限状態変換器を、確実に作成することができる。

【0 0 0 9】

また、請求項 5 に記載のコンピュータ読み取り可能な記録媒体は、請求項 4 に記載の有限状態変換器作成プログラムを記録している。

従って、コンピュータによって、請求項 5 に記載のコンピュータ読み取り可能な記録媒体から請求項 4 に記載の有限状態変換器作成プログラムを読み取って実

行することにより、文法規則の適用頻度に関する統計情報に基づく弧の置き換え優先度が高い弧から順に置き換え操作が適用されるので、限られた大きさで、より多くの文を解析可能な有限状態変換器を、確実に作成することができる。

【0010】

また、請求項6に記載の有限状態変換器作成方法は、漸進的構文解析に用いる有限状態変換器を作成する方法であって、文脈自由文法に基づく文法規則の集合を表すネットワークの集合であると共に、前記各ネットワークにおける非終端記号による遷移が他のネットワークによって定義される再帰的構造を有する再帰遷移ネットワークを作成する再帰遷移ネットワーク作成ステップと、開始記号を入力ラベルとする弧を持つ有限状態変換器を初期の有限状態変換器とし、前記有限状態変換器の弧をその入力ラベルに対応した前記再帰遷移ネットワーク中のネットワークで置き換え、さらに、その置き換えによって新たに作成された弧を、前記再帰遷移ネットワーク中の別のネットワークに置き換える操作を再帰的に繰り返す弧置き換えステップと、文法規則の適用頻度に関する統計情報に基づいて、前記有限状態変換器における入力ラベルが非終端記号である全ての弧について各々に対応する構文木の節点の導出確率を計算し、得られた導出確率を弧の置き換え優先度とする優先度計算ステップと、を備え、前記弧置き換えステップにおいて、前記弧の置き換え優先度が高い弧から順に置き換え操作を適用すると共に、前記置き換え操作が繰り返し適用されることによって前記有限状態変換器が所定の大きさに達したときに前記弧の置き換え操作の適用を終了することを特徴とする。

従って、請求項6に記載の有限状態変換器作成方法によれば、文法規則の適用頻度に関する統計情報に基づく弧の置き換え優先度が高い弧から順に置き換え操作が適用されるので、限られた大きさで、より多くの文を解析可能な有限状態変換器を、確実に作成することができる。

【0011】

また、請求項7に記載の有限状態変換器作成方法は、前記弧置き換えステップにおける前記弧の置き換え操作の適用が終了した後、非終端記号を入力ラベルに持つ弧を除去しつつ、弧の置き換え操作の適用をさらに実行する弧除去ステップ

、を備えたことを特徴とする。

従って、請求項 7 に記載の有限状態変換器作成方法によれば、構文解析時に使用されない非終端記号を入力ラベルに持つ弧を除去しつつ、さらに弧の置き換えを行うので、より一層多くの文を解析可能な有限状態変換器を確実に作成することができる。

【0012】

また、請求項 8 に記載の有限状態変換器作成方法は、前記節点の導出確率が、構文木における開始記号から対象の節点までのパス上の各節点について順に文法規則が適用される確率であることを特徴とする。

従って、請求項 8 に記載の有限状態変換器作成装置によれば、構文木における開始記号から対象の節点までのパス上の各節点について順に文法規則が適用される確率を弧の置き換え優先度として用いて弧の置き換え操作を行うことによって、より多くの文を解析可能な有限状態変換器を確実に作成することができる。

【0013】

また、請求項 9 に記載の漸進的構文解析装置は、漸進的に構文解析を行うように構成された構文解析装置であって、請求項 6 乃至 8 のいずれかに記載の方法によって作成された有限状態変換器と、その有限状態変換器へ単語を入力する度に状態遷移に伴って出力される構文木を順次接続する接続処理手段と、を備えたことを特徴とする。

従って、請求項 9 に記載の漸進的構文解析装置によれば、請求項 6 乃至 8 のいずれかに記載の方法によって作成された有限状態変換器、すなわち、文法規則の適用頻度に関する統計情報に基づく弧の置き換え優先度が高い弧から順に置き換え操作が適用された有限状態変換器を備え、接続処理手段が、その有限状態変換器へ単語を入力する度に状態遷移に伴って出力される構文木を順次接続するように構成されているので、文脈自由文法を近似変換した限られた大きさの有限状態変換器を用いて、より多くの文について漸進的に構文解析を行うことができる。

【0014】

【発明の実施の形態】

以下、本発明を具体化した有限状態変換器作成装置、プログラム、記録媒体、

作成方法、及び漸進的構文解析装置の一実施形態について図面を参照しつつ説明する。

まず、本実施形態の有限状態変換器作成装置 1 の全体構成について、図 1 を参照しつつ説明する。

有限状態変換器作成装置 1 は、図 1 に示すように、再帰遷移ネットワーク作成部 2 と、弧置き換え部 3 と、優先度計算部 4 と、弧除去部 5 とから構成され、文法規則の適用頻度に関する統計情報を記憶する統計情報記憶装置 11 が接続されている。但し、後述する弧の除去処理を行わない場合は、弧除去部 5 を省略して構成することも可能である。

【0015】

有限状態変換器作成装置 1 は、具体的には、CPU、ROM、RAM、ハードディスク装置、CD-ROM 装置等を備えたコンピュータによって実現され、例えば、コンピュータを再帰遷移ネットワーク作成部 2、弧置き換え部 3、優先度計算部 4、及び弧除去部 5 として機能させるための有限状態変換器作成プログラムをハードディスク装置に記憶させ、CPU がハードディスク装置から有限状態変換器作成プログラムを読み出して実行するように構成される。また、CD-ROM 等に記録された文法規則の適用頻度に関する統計情報が予め CD-ROM 装置等を介して予めコンピュータに読み込まれてハードディスク装置に記憶されている場合は、ハードディスク装置が統計情報記憶装置 11 として機能する。尚、文法規則の適用頻度に関する統計情報としては、例えば、構文木付き ATR 音声言語データベース(日本語対話)を用いることができる。

尚、再帰遷移ネットワーク作成部 2 が本発明の再帰遷移ネットワーク作成手段を、弧置き換え部 3 が弧置き換え手段を、優先度計算部 4 が優先度計算手段を、弧除去部 5 が弧除去手段をそれぞれ構成するものである。また、再帰遷移ネットワーク作成部 2 における処理内容が本発明の再帰遷移ネットワーク作成ステップに、弧置き換え部 3 における処理内容が弧置き換えステップに、優先度計算部 4 における処理内容が優先度計算ステップに、弧除去部 5 における処理内容が弧除去ステップにそれぞれ相当するものである。

【0016】

次に、有限状態変換器作成装置 1 を構成する上述した各部の処理内容について図面を参照しつつ説明する。

まず、有限状態変換器作成装置 1 各部の処理内容の説明に先立って、有限オートマトン、有限状態変換器、文脈自由文法をそれぞれ定義する。

はじめに、有限オートマトンを定義する。有限オートマトンは、5 項組 (Σ, Q, q_0, F, E) で定義される。 Σ はアルファベットの有限集合、 Q は状態の有限集合、 $q_0 \in Q$ は初期状態、 $F \subseteq Q$ は最終状態の集合、 E は弧の有限集合である。また、 $E \subseteq Q \times \Sigma \times Q$ である。

有限オートマトンは、1 つの初期状態と 1 つ以上の最終状態を持ち、弧のラベルに従って、状態を遷移するネットワークである。また、弧 $(p, A, q) \in E$ ($p, q \in Q, A \in \Sigma$) に対して、状態 p を弧の始点、状態 q を弧の終点と呼ぶ。

【0017】

次に、有限状態変換器を定義する。有限状態変換器は、6 項組 $(\Sigma_I, \Sigma_O, Q, q_0, F, E)$ で定義される。 Σ_I, Σ_O は、それぞれ入力アルファベット、出力アルファベットの有限集合、 Q は状態の有限集合、 $q_0 \in Q$ は初期状態、 $F \subseteq Q$ は最終状態の有限集合、 E は弧の有限集合である。ただし、 $E \subseteq Q \times \Sigma_I \times \Sigma_O \times Q$ である。

有限オートマトンでは弧に入力ラベルを割り当てたが、さらに出力ラベルを割り当てたものが有限状態変換器である。有限状態変換器では、 Σ_I の要素が入力されたときに、 Σ_O の要素を出力して遷移する。有限状態変換器を用いることによって、システムに入力された記号列の受理だけでなく、入力に対応する記号列の出力が可能となる。

【0018】

最後に文脈自由文法 G を定義する。 G は、4 項組 (N, T, P, S_0) で定義される。 N, T はそれぞれ非終端記号、終端記号の有限集合である。 $S_0 \in N$ は開始記号であり、この文法から生成される構文木の根節点となる。さらに、 P は文法規則の集合である。各規則は、 $A \rightarrow \alpha$ ($A \in N, \alpha = (N \cup T)^+$) の形式で書かれ、 A が α に書き換えられることを示す。自然言語の構造の多くは、文

脈自由文法で記述可能である。

【0019】

次に、有限状態変換器作成装置 1 を構成する各部の処理について説明する。本実施形態では、文脈自由文法を再帰遷移ネットワークで表現し、得られた再帰遷移ネットワークの中の弧を、別のネットワークで置き換えていくことにより、有限状態変換器を獲得する。以下では、まず、再帰遷移ネットワーク作成部 2 によって実行される再帰遷移ネットワークの作成処理について述べ、次に、弧置き換え部 3、優先度計算部 4、及び弧除去部 5 における再帰遷移ネットワークによる置き換え操作を利用した有限状態変換器の作成処理について述べる。

【0020】

(再帰遷移ネットワーク作成部 2 における再帰遷移ネットワーク作成処理)

再帰遷移ネットワークは、非終端記号による遷移を許したネットワークの集合である。再帰遷移ネットワークにおいて非終端記号による遷移は、他のネットワークによって定義されるという再帰的な構造を持っている。再帰遷移ネットワークと文脈自由文法の解析能力は等価である。以下では、文脈自由文法から、それと等価な再帰遷移ネットワークを作成する方法を述べる。

まず、範疇 X に対して、左辺が範疇 X である文法規則の集合 P_X を表現するネットワーク M_X は、以下のように定義される。ネットワーク M_X は、5 項組 $(\Sigma, Q_X, i_X, F_X, E_X)$ である。ただし、 $\Sigma = T \cup N$ 、 i_X は初期状態、 F_X は最終状態の集合である。このとき、 $F_X = \{f_X\}$ とする。また、 Q_X は、状態の有限集合、 E_X は弧の有限集合である。

Q_X の要素を表現するために、ドット記号 (\cdot) 付き文法規則を導入する。ドット記号付き文法規則は、 $X \rightarrow \alpha \cdot \beta$ のように、ドット記号を文法規則の右辺の任意の位置に挿入したものである。さらに、表記の簡単化のために、ドット記号付き規則を、その左辺、右辺のドット記号の左側、ドット記号の右側の 3 項組で表現する。例えば、 $X \rightarrow \alpha \cdot \beta$ は、 (X, α, β) と表現する。この表現を用いると、 Q_X は次の数式 1 で表される集合である。

【数 1】

$$Q_X = \{(X, \alpha, \beta) \mid X \rightarrow \alpha\beta \in P_X, \alpha, \beta \in (N \cup T)^+\} \\ \cup \{i_X, f_X\}$$

また、 E_X は、次の数式 2 で表される集合である。

【数 2】

$$E_X = \{((X, \alpha, A\beta), A, (X, \alpha A, \beta)) \\ \mid X \rightarrow \alpha A\beta \in P_X\} \\ \cup \{(i_X, A, (X, A, \beta)) \mid X \rightarrow A\beta \in P_X\} \\ \cup \{((X, \alpha, A), A, f_X) \mid X \rightarrow \alpha A \in P_X\} \\ \cup \{(i_X, A, f_X) \mid X \rightarrow A \in P_X\}$$

ただし、 $X \in N$ 、 $A \in N \cup T$ 、 $\alpha, \beta \in (N \cup T)^+$ である。

例えば、 P_X が図 2 に挙げた規則の集合であるとき、 M_X は図 3 に示すネットワークである。 M_X の初期状態 i_X から最終状態 f_X へのパスは、 P_X の中の 1 つの文法規則に対応する。従って、文法規則の右辺の記号列を M_X に入力すれば、文法規則に対応する M_X のパス上を通り、 i_X から f_X へと遷移できる。本実施形態の手法では、再帰遷移ネットワーク M を、 M_X の集合として数式 3 により定義する。

【数 3】

$$\mathcal{M} = \{M_X \mid X \in N\}$$

【0021】

(再帰遷移ネットワーク作成部 2 における再帰遷移ネットワークの簡単化処理)

上述した処理によって作られる再帰遷移ネットワークには、始点が等しく、かつ、同じラベルを持つ弧が複数存在するため、冗長性を持ち、決定的に遷移できない。そのため、有限オートマトンの最小化手法に基づき、状態を統合する。すなわち、再帰遷移ネットワークの各 M_X ($X \in N$) について、等価に変換可能であれば、状態を統合する。ただし、 F_X の要素数を 2 個以上にする状態の統合は

認めない。 M_X を置き換え操作に用いる際に、置き換え操作が容易に行えるようにするためである。

M_X の簡単化は、表1に示す手順に従って、状態を統合することにより実現する。まず、 M_X に変化がなくなるまで手順1の操作を繰り返して状態を統合し、次に手順2の操作を M_X に変化がなくなるまで繰り返す。以下の手続き中の記号は、それぞれ $q, q', q'' \in Q_X, A \in \Sigma_I$ である。

【表1】

ネットワーク M_X の簡単化

手順 1	ある q が存在し, $(q, A, q') \in E_X$ かつ $(q, A, q'') \in E_X$ かつ $q', q'' \notin F_X$ ならば, q' と q'' を統合する.
手順 2	状態 q', q'' について, 任意の記号 $A \in \Sigma_I$ に対して, $((q', A, q) \in E_X$ かつ $(q'', A, q) \in E_X)$, または, $((q', A, q) \notin E_X$ かつ $(q'', A, q) \notin E_X)$ となる q が存在するならば, q' と q'' を統合する.

図4に、上述した統合操作の一例を示す。手順1では、同じ状態からAで遷移する状態を統合する。手順2では、Dで遷移する先の状態が等しく、他の記号による遷移先を持たない2つの状態を統合する。簡単化された再帰遷移ネットワークでは、ある状態から同じラベルで遷移できる状態は、最大でも、最終状態とそれ以外の状態、それぞれ1つずつである。

【0022】

(弧置き換え部3における再帰遷移ネットワークを用いた有限状態変換器の作成処理)

次に、上述した再帰遷移ネットワーク作成処理によって作成された再帰遷移ネットワークを用いた有限状態変換器の作成処理について述べる。まずはじめに、初期有限状態変換器 M_0 を数式4により定義する。

【数4】

$$M_0 = (Q_0, \Sigma_I, \Sigma_O, i, F, E_0)$$

各記号は、それぞれ $Q_0 = \{i, f\}$, $\Sigma_I = N \cup T$, $\Sigma_O \subset (([N])^* ($

$\Sigma_I) * (N)] *)$, $F = \{f\}$, $E_0 = \{(i, S_0, S_0, f)\}$ である。

初期有限状態変換器 M_0 を表した図が、図 5 である。 M_0 の弧をネットワーク M_{S_0} で置き換え、さらに、新たに作られた弧に対する置き換え操作を再帰的に繰り返すことによって、有限状態変換器を獲得する。置き換え操作は、入力ラベルが非終端記号である弧に対して行い、 X を入力ラベルとして持つ弧は、 M_X で置き換えられる。

【0023】

次に、置き換え操作の前後における、有限状態変換器の変化について述べる。有限状態変換器 M_0 に対して、何回かの置き換え操作を実行して得られた有限状態変換器を M_j とする。 M_j を $(Q_j, \Sigma_I, \Sigma_O, i, F, E_j)$ とする。弧 $e = (q_s, X, o_l X o_r, q_e) \in E_j$ を M_X で置き換えて得られる有限状態変換器を M_j とする。ただし、 o_l 、 o_r は、それぞれ出力アルファベット中の、左括弧付範疇の系列 $([N])^*$ 及び右括弧付範疇の系列 $(N)]^*$ を表す。 M'_j は、 Q_j と弧 E_j とに、新たに状態と弧とが追加されて作成される。従って、状態の集合と弧の集合とが変化するため、 M'_j を $(Q'_j, \Sigma_I, \Sigma_O, i, F, E'_j)$ とする。このとき、 Q'_j 、 E'_j は数式 5、数式 6 のように作ることができる。ただし、 $q_1 \neq i_X$, $q_2 \neq f_X$ である。

【数 5】

$$Q'_j = Q_j \cup \{eq \mid q \in (Q_X - \{i_X, f_X\})\}$$

【数 6】

$$\begin{aligned} E'_j &= (E_j - \{e\}) \\ &\cup \{(eq_1, A, A, eq_2) \mid (q_1, A, q_2) \in E_X\} \\ &\cup \{(q_s, A, o_l[XA, eq_2) \mid (i_X, A, q_2) \in E_X\} \\ &\cup \{(eq_1, A, A_X]o_r, q_e) \mid (q_1, A, f_X) \in E_X\} \\ &\cup \{(q_s, A, o_l[XA_X]o_r, q_e) \mid (i_X, A, f_X) \in E_X\} \end{aligned}$$

【0024】

置き換え操作の例を図 6 に示す。尚、図 6 において、 S_0 (開始記号)、 S (

文)、P (後置詞)、PP (後置詞句)、NP (名詞句)、V (動詞)、VP (動詞句)、\$ (終止符)である。図6の左の図は、PPを入力ラベルに持つ弧を、左辺がPPである文法規則を表現するネットワーク M_{PP} で置き換える操作を示しており、右の図は、対応する構文木を表している。

置き換え操作は一般に無限に続けることができる。しかし、有限状態変換器作成装置が実現されるコンピュータのメモリ領域は有限であり、作成できる有限状態変換器の大きさには限りがある。そこで、本実施形態では、有限状態変換器の大きさを表している弧の数に関して閾値を設定し、弧の数が閾値 λ に達したとき(すなわち、弧の置き換え操作の繰り返しによって有限状態変換器が所定の大きさに達したとき)に弧の置き換え操作を終了することによって、有限状態変換器の作成を近似的に実現する。

【0025】

(優先度計算部4における統計情報を利用した弧の置き換え順序決定処理)

上述した弧置き換え部3によって実行される弧の置き換え処理によって、漸進的構文解析に用いる有限状態変換器を作成できる。しかし、単純に置き換え操作を繰り返すだけでは、本当に必要な弧を置き換える前に、置き換え操作が打ち切られてしまう可能性がある。従って、置き換え操作を実行するときには、置き換える弧の選択が重要となる。優先度計算部4では、統計情報記憶装置11に記憶された文法規則の適用頻度に関する統計情報を用い、有限状態変換器の弧と構文木の節点との対応関係を利用して、節点の導出確率が高い節点に対応する弧ほど置き換えるの必要性が高いと判断し、弧の置き換え順序を決定する。

まず、有限状態変換器の弧と構文木の節点の対応関係について説明する。有限状態変換器の弧は、 S_0 を入力ラベルとする弧からネットワークによる置き換え操作を再帰的に実行していくことにより作成される。ネットワークは文法規則の集合を表現しているため、文法規則を適用していると考えられることもできる。一方、文脈自由文法において、トップダウンに構文木を生成する場合にも、はじめに S_0 に対して文法規則を適用し、作られた節点に対して文法規則を再帰的に適用することによって、節点は生成される。すなわち、弧と節点は、共に開始記号から文法規則を再帰的に適用して作られるものである。これらの適用操作は対応づ

けることができ、その操作によって作られた弧と節点も対応づけて考えることができる。図6に、弧と節点の対応の例を番号を用いて示す。例えば、図中の1で示される弧と節点とは、開始記号 S_0 に対して、ともに $S_0 \rightarrow S \$$, $S \rightarrow \dots VP$, $VP \rightarrow PPV$ の順に規則が適用されて作られるため、対応する。

【0026】

有限状態変換器を用いた構文解析において、ある節点を含む構文木を生成するためには、その節点に対応する弧が置き換えられなければならない。しかし、作成できる弧の数は有限であるため、最終的に、全ての弧が置き換えられるわけではない。つまり、全ての構文木が生成できるわけではなく、その中で、できるだけ多くの構文木を生成できる有限状態変換器を作成するためには、弧の置き換え順序を考慮する必要がある。弧の置き換え順序を決定するための指標を、置き換え優先度と呼ぶことにする。導出確率の高い節点を含む構文木ほど頻繁に生成されるため、その節点に対応する弧は、優先して置き換える必要があると考えられる。そこで、置き換え優先度の値を、対応する節点の導出確率とする。有限状態変換器の作成では、統計情報記憶装置11に記憶された文法規則の適用頻度に関する統計情報を用いて、入力ラベルが非終端記号である全ての弧に対して置き換え優先度を計算し、その値が高い弧から順に弧置き換え部3による置き換え操作を適用する。

【0027】

次に、節点の導出確率の計算方法について述べる。構文木の節点は、 S_0 からその節点までのパス上の節点に、文法規則が順次適用されて作られる。そこで、節点の導出確率を、 S_0 から導出確率を求めたい節点までのパス上の各節点に、順に文法規則が適用される確率とする。図7では、節点 $X_{rM}(lM)$ は、構文木の根節点 S_0 に対して文法規則 r_1 が適用され、 r_1 が生成した節点の中で左から l_1 番目の節点 $X_{r1}(l_1)$ に文法規則 r_2 が適用され、最後に、文法規則 r_{M-1} が生成した節点の左から l_{M-1} 番目の節点に文法規則 r_M が適用されて作られる。この節点の導出確率 $P(X_{rM}(lM))$ を、数式7で計算する。

【数 7】

$$\begin{aligned}
P(X_{r_M(l_M)}) &= P(r_{1(l_1)}, r_{2(l_2)}, \dots, r_{M-1(l_{M-1})}, r_{M(l_M)}) \\
&= P(r_{1(l_1)}) \\
&\times P(r_{2(l_2)} \mid r_{1(l_1)}) \\
&\times P(r_{3(l_3)} \mid r_{1(l_1)}, r_{2(l_2)}) \\
&\vdots \\
&\times P(r_{M(l_M)} \mid r_{1(l_1)}, \dots, r_{M-1(l_{M-1})})
\end{aligned}$$

$r_i(l_i)$ は、文法規則 r_i が適用され、かつ、次に適用される文法規則 r_{i+1} が、 r_i の右辺 l_i 番目の要素が生成する節点に適用されることを示す。このとき、文法規則が適用される位置を考えるのは、同じ範疇であっても、位置によって適用されやすい規則は異なるためである。例えば、文法規則 $N \rightarrow NN$ に対して、右辺の 1 番目の N と 2 番目の N とでは、適用されやすい文法規則は異なる。

【0028】

ここで、数式 7 中の $P(r_i(l_i) \mid r_{1(l_1)}, \dots, r_{i-1}(l_{i-1}))$ の値は、次の文法規則の適用位置にかかわらないため、数式 7 は、数式 8 とすることができる。

【数 8】

$$\begin{aligned}
P(X_{r_M(l_M)}) &= P(r_{1(l_1)}, r_{2(l_2)}, \dots, r_{M-1(l_{M-1})}, r_M) \\
&= P(r_1) \\
&\times P(r_2 \mid r_{1(l_1)}) \\
&\times P(r_3 \mid r_{1(l_1)}, r_{2(l_2)}) \\
&\vdots \\
&\times P(r_M \mid r_{1(l_1)}, \dots, r_{M-1(l_{M-1})})
\end{aligned}$$

このようにして、節点の導出確率は求められる。しかし、数式 8 のように、節点の導出において適用された全ての文法規則を条件として文法規則の適用確率を求めると、スパースネス問題が発生し、作成する有限状態変換器が学習データに依存したものとなる。そこで、優先度計算部 4 では、ある節点に対して文法規則が適用される確率は、その節点から順にさかのぼって最初に到達する $N-1$ 個の

節点を生成した文法規則とその適用位置だけに依存するものとする。また、得られた適用確率に対して、低次の条件付き適用確率と線形補間を行うことによって、スムージングを行う。

【0029】

まず、数式 9 に示される近似した文法規則の適用確率 P の計算方法について述べる。

【数 9】

$$P(r_i | r_{i-N+1(l_{i-N+1})}, \dots, r_{i-1(l_{i-1})})$$

ある節点に対して文法規則を適用するとき、その節点から S_0 までのパス上を順にさかのぼっていき、適用された文法規則と、その右辺の中で次の規則が適用された位置をペアとする $N-1$ 項組を獲得する。これに、今、適用する文法規則を合わせることによって、 $(r_1(l_1), \dots, r_{N-1}(l_{N-1}), r_N)$ の N 項組で表すことができる。例えば、図 8 では、6 つの文法規則が適用されて構文木が作られている。この構文木からは 6 つの組が得ることができ、例えば $N=3$ のときには、図 8 に示される 6 つの 3 項組を獲得できる。ただし、構文木の開始記号より上の位置では、ヌル規則 ‘#’ が適用されていると仮定している。

【0030】

学習データから獲得した N 項組の集合を用いて、 $r_1(l_1), \dots, r_{N-1}(l_{N-1})$ を条件とした文法規則 r_N の適用確率を、数式 10 で計算する。ただし、 $C(X)$ は、 X の出現回数を示す。

【数 10】

$$P(r_N | r_1(l_1), \dots, r_{N-1}(l_{N-1})) \\ = \frac{C(r_1(l_1), \dots, r_{N-1}(l_{N-1}), r_N)}{\sum_{r_N} C(r_1(l_1), \dots, r_{N-1}(l_{N-1}), r_N)}$$

さらに、文法規則の適用確率には、数式 11 によって線形補間した値を用いる。ただし、 $\lambda_1, \dots, \lambda_N$ は補間係数である。

【数 1 1】

$$\begin{aligned}
& \hat{P}_N(r_N | r_{1(l_1)}, \dots, r_{N-1(l_{N-1})}) \\
& = \lambda_N P_N(r_N | r_{1(l_1)}, \dots, r_{N-1(l_{N-1})}) \\
& + \lambda_{N-1} P_{N-1}(r_N | r_{2(l_2)}, \dots, r_{N-1(l_{N-1})}) \\
& \vdots \\
& + \lambda_2 P_2(r_N | r_{N-1(l_{N-1})}) \\
& + \lambda_1 P_1(r_N | LHS(r_N))
\end{aligned}$$

ただし、LHS (r_N) は r_N の左辺範疇を表す。P₁ (r_N | LHS (r_N)) 以外の条件に LHS (r_N) を含めないのは、文法規則 r_{N-1} の位置 l_{N-1} にある範疇は、LHS (r_N) であるとはわかるためである。

最終的に、本手法では数式 1 2 を用いて節点の導出確率を求める。

【数 1 2】

$$\begin{aligned}
& P(X_{r_{M(l_M)}}) \\
& = \prod_{i=1}^M \hat{P}(r_i | r_{i-N+1(l_{i-N+1})}, \dots, r_{i-1(l_{i-1})})
\end{aligned}$$

ただし、再帰遷移ネットワークの状態を統合した影響により、複数の文法規則から作られている弧が再帰遷移ネットワークには存在する。そのため、1つの弧に対して構文本の複数の節点に対応することがあるが、その場合には、対応する全ての節点の導出確率の和が節点の導出確率であるとする。

【0031】

(弧除去部 1 4 における非終端記号をラベルに持つ弧の除去処理)

先に述べた弧置き換え部 3 によって実行される有限状態変換器作成処理では、弧の数が閾値 λ に達したら、すぐに置き換え操作を打ち切るため、ネットワークで置き換えられなかった非終端記号を入力ラベルに持つ弧はそのまま有限状態変換器中に残される。しかし、本実施形態の解析手法では、弧の入力ラベルとシステムに入力される単語の品詞が一致する場合にのみ遷移するため、非終端記号を入力ラベルに持つ弧は解析時には使用されない。従って、これらの弧をそのまま残しておくことは無駄であり、弧を除去しても問題とならない。それどころか、これらの弧を除去しつつ、さらに弧を置き換えることができれば、有限状態変換

器の解析能力の向上が期待できる。以下、非終端記号をラベルに持つ弧を除去しつつ、さらに置き換え操作を継続する処理について述べる。

まず、弧置き換え部 3 による処理により有限状態変換器を作成する。弧の数が、閾値 λ に達して置き換え操作の適用が停止したのち、以下のアルゴリズムを実行する。

【0032】

(非終端記号を入力ラベルとする弧の除去手続き)

1. 非終端記号のラベルの中で最も置き換え優先度の高い弧 e を、次に置き換える弧として選択する。ここで、弧 e の入力ラベルを $I(e)$ とする。
2. e の置き換える有効性をチェックする。有効でないときには e を除去し、1. へ戻る。
3. 有限状態変換器の中で、非終端記号を入力ラベルに持つ弧を、置き換え優先度の低い順に除去する。除去する弧の数は、 $\lambda - ((\text{有限状態変換器の弧の数}) - (M_{I(e)} \text{ が持つ弧の数}) - 1)$ 個である。ただし、この値が負である場合には除去しない。
4. 弧 e をネットワーク $M_{I(e)}$ で置き換える。
5. 有限状態変換器に非終端記号を入力ラベルとする弧が残っていれば、再び 1. から処理を繰り返す。

上記のアルゴリズムの 2. の有効性のチェックでは、弧 e について、弧 e の始点の状態を遷移先とする弧が存在するか、もしくはその状態が初期状態であるかをチェックし、さらに、弧 e の終点の状態を遷移元とする弧が存在するか、もしくはその状態が最終状態であるかをチェックする。どちらか一方でも当てはまらなければ、弧 e は解析に使われないため除去される。

この操作によって、残された弧の中で、置き換え優先度の高い弧はさらに置き換えられ、置き換え優先度の低い弧は除去される。しかし、弧を除去することによって、初期状態から到達できない弧や、最終状態まで到達できない弧が新たに現れる。これらの弧も解析に用いることはできない。従って、弧を除去するときには、その影響について調査し、使用できない弧がさらに出現するときにはその弧もまとめて除去する。従って、弧を除去するときには以下の操作を行う。

【0033】

(不要な弧の除去方法)

弧を除去する場合に、その弧の始点、終点の状態を共有している弧について、以下の点をチェックする。もしどれか1つに該当すれば、その指示に従って弧を除去し、さらに除去した弧について再帰的に同じ操作を実行する。

(1) 除去した弧の始点を遷移先とする弧が存在しない場合、その状態を始点とする全ての弧を除去する。

(2) 除去した弧の始点を遷移元とする弧が他に存在しない場合、その状態を終点とする全ての弧を除去する。

(3) 除去した弧の終点を遷移先とする弧が他に存在しない場合、その状態を始点とする全ての弧を除去する。

(4) 除去した弧の終点を遷移元とする弧が存在しない場合、その状態を終点とする全ての弧を除去する。

(1) から (4) までの操作を図にまとめると、図9のようになる。図9の点線で示された弧は、それぞれのパターンにおいて存在しない弧を示す。いずれの図でも、中央の×印の弧が除去されたときに、点線の弧がないために、さらに除去される弧が×印で示されている。

以上詳述した有限状態変換器作成装置1における再帰遷移ネットワーク作成部2、弧置き換え部3、優先度計算部4、及び弧除去部5における各処理ステップが実行された結果として、漸進的構文解析に用いる有限状態変換器が獲得される。

【0034】

(漸進的構文解析装置21による漸進的な構文木生成)

次に、上述した有限状態変換器作成装置1によって作成された有限状態変換器22を用いた漸進的構文解析装置21について、図面を参照しつつ説明する。

漸進的構文解析装置21は、図10に示すように、入力装置31と、有限状態変換器22と、接続処理部23と、出力装置32とから構成されている。漸進的構文解析装置21は、具体的には、CPU、ROM、RAM、ハードディスク装置、音声入力装置、ディスプレイ装置等を備えたコンピュータによって実現され

る。また、接続処理部 23 が、本発明の接続処理手段を構成するものである。

【0035】

入力装置 31 は、構文解析の対象となる文を入力するための装置であり、具体的には、音声入力装置、キーボード等の入力装置によって構成される。入力装置 31 は、外部から入力された文（単語列）を、順次、有限状態変換器 22 に入力する。

有限状態変換器 22 は、文法規則の適用の過程を予め計算した結果を有限状態変換器として表現したものであって、上述した有限状態変換器作成装置 1 によって作成されたものである。有限状態変換器 22 は、入力装置 31 によって入力される単語列に対して状態遷移すると共に文法規則適用により生成される構文木を順に出力する。有限状態変換器 22 は、具体的には、ROM 又はハードディスク装置に記憶された有限状態変換器プログラムを CPU が読み出して実行することにより実現される。

接続処理部 23 は、有限状態変換器 22 によって出力された構文木を順次接続する。従って、文の途中段階でも、それまでの入力に対する構文木を生成することができる。接続処理部 23 は、具体的には、ROM 又はハードディスク装置に記憶された接続処理プログラムを CPU が読み出して実行することにより実現される。

出力装置 32 は、有限状態変換器 22 及び接続処理部 23 によって生成された構文解析結果としての構文木を出力する。出力部 32 は、具体的には、構文解析結果をディスプレイ装置による表示として、RAM 又はハードディスク上へのファイル等として出力する。

【0036】

次に、漸進的構文解析装置 21 において漸進的に構文木を生成する処理の詳細内容について説明する。本実施形態の漸進的構文解析装置 21 では、基本的には、入力装置 31 から有限状態変換器 22 へ単語をつぎつぎと入力することによって、状態を遷移して、構文木の出力を得ることができる。しかし、上述した有限状態変換器作成装置 1 によって作られる有限状態変換器 22 は非決定性であるため、ある入力に対して、複数の遷移先が存在する可能性がある。漸進的構文解析

では、入力に合わせて構文構造を出力するべきであると考え、本実施形態では幅優先探索を行い、構文木を出力する。すなわち、現在の状態と、これまでに出力された構文木とを表現する記号列のペアを要素とするリストを持ち、1単語ずつが入力されるたびに、現在の状態から遷移できる全ての状態に状態遷移する。そのとき、接続処理部23が、それ以前に入力された単語列に対する出力構文木を示す記号列に、遷移した弧に記述された出力ラベルを接続して、新しい構文木を生成する。

【0037】

漸進的構文解析装置21における動作例を図11に示す。尚、図11において示される各出力記号が表す意味内容を以下に括弧書きにて示す。すなわち、S0（開始記号）、S（文）、NP（名詞句）、N-HUTU（普通名詞句）、HUTU-MEISI（普通名詞）、VAUX（動詞句）、VERB（動詞）、AUX（助詞）、AUX-DE（助詞「で」）、AUXSTEM（助詞語幹）、AUXSTEM-MASU（助詞語幹「（ござい）ます」）、INFL（活用語尾）、INFL-SPE-SU（活用語尾「す」）、\$（句点）である。

入力装置31より有限状態変換器22に1単語入力されるごとに有限状態変換器22が状態遷移し、遷移した弧の出力ラベルが接続処理部23によって接続される。ここで、出力記号列（接続された複数の出力ラベル）は1つの構文木を表している。例えば、品詞‘HUTU-MEISI’（普通名詞）が入力されたときの出力記号列は、図12の左側に示された構文木を表しており、‘AUX-DE’（助詞「で」）まで入力されたときの出力記号列は、図12の右側に示された構文木を表している。このように、単語が入力されるごとに、次々に構文木を拡張していく。この例では、遷移に曖昧性を含んでいないため、各品詞の入力に対して構文木は一つしか出力されていないが、前に述べたように、複数の状態に遷移可能であれば、その数だけ、状態と記号列のペアは保持され、構文木が作られる。

【0038】

以上詳述したことから明らかなように、本実施形態によれば、有限状態変換器作成装置1は、文脈自由文法に基づく文法規則の集合を表すネットワークの集合であると共に、前記各ネットワークにおける非終端記号による遷移が他のネットワークによって定義される再帰的構造を有する再帰遷移ネットワークを作成する

再帰遷移ネットワーク作成部 2 と、開始記号を入力ラベルとする弧を持つ有限状態変換器を初期の有限状態変換器とし、前記有限状態変換器の弧をその入力ラベルに対応した前記再帰遷移ネットワーク中のネットワークで置き換え、さらに、その置き換えによって新たに作成された弧を、前記再帰遷移ネットワーク中の別のネットワークに置き換える操作を再帰的に繰り返す弧置き換え部 3 と、文法規則の適用頻度に関する統計情報に基づいて、前記有限状態変換器における入力ラベルが非終端記号である全ての弧について各々に対応する構文木の節点の導出確率を計算し、得られた導出確率を弧の置き換え優先度とする優先度計算部 4 と、を備え、前記弧置き換え部 3 は、前記優先度計算部 4 で求められた前記弧の置き換え優先度が高い弧から順に置き換え操作を適用すると共に、前記置き換え操作が繰り返し適用されることによって前記有限状態変換器が所定の大きさに達したときに前記弧の置き換え操作の適用を終了することを特徴とする。

【0039】

従って、有限状態変換器作成装置 1 によれば、文法規則の適用頻度に関する統計情報に基づく弧の置き換え優先度が高い弧から順に置き換え操作が適用されるので、限られた大きさで、より多くの文を解析可能な有限状態変換器を、確実に作成することができる。

また、本実施形態によれば、有限状態変換器作成装置 1 は、前記有限状態変換器が所定の大きさに達したことにより前記弧置き換え部 3 による前記弧の置き換え操作の適用が終了した後、非終端記号を入力ラベルに持つ弧を除去しつつ、弧の置き換え操作の適用をさらに実行する弧除去手段 5、を備えており、構文解析時に使用されない非終端記号を入力ラベルに持つ弧を除去しつつ、さらに弧の置き換えを行うので、より一層多くの文を解析可能な有限状態変換器を確実に作成することができる。

また、本実施形態によれば、有限状態変換器作成装置 1 は、構文木における開始記号から対象の節点までのパス上の各節点について順に文法規則が適用される確率を弧の置き換え優先度として用いて弧の置き換え操作を行うことによって、より多くの文を解析可能な有限状態変換器を確実に作成することができる。

【0040】

また、本実施形態によれば、漸進的構文解析装置 2 1 は、有限状態変換器作成装置 1 によって作成された有限状態変換器 2 2 と、その有限状態変換器 2 2 へ単語を入力する度に状態遷移に伴って出力される構文木を順次接続する接続処理部 2 3 と、を備えたことを特徴とする。

従って、漸進的構文解析装置 2 1 によれば、文法規則の適用頻度に関する統計情報に基づく弧の置き換え優先度が高い弧から順に置き換え操作が適用された有限状態変換器 2 2 を備え、接続処理部 2 3 が、その有限状態変換器 2 2 へ単語を入力する度に状態遷移に伴って出力される構文木を順次接続するように構成されているので、文脈自由文法を近似変換した限られた大きさの有限状態変換器 2 2 を用いて、より多くの文について漸進的に構文解析を行うことができる。

【0 0 4 1】

尚、本発明は上述した各実施の形態に限定されるものではなく、本発明の主旨を逸脱しない範囲で種々の変更を施すことが可能である。

例えば、前記実施形態では、漸進的構文解析装置 2 1 を単体で用いる例を示したが、漸進的構文解析装置 2 1 を同時通訳システム又は音声認識システムの一部として組み込むことにより、実時間性に優れ且つ正解率の高い同時通訳システム又は音声認識システムを実現することができる。また、漸進的構文解析装置 2 1 を組み込んだ音声認識システムをロボットに搭載することにより、極めて応答性に優れた音声入力ロボットや対話型ロボットを実現することができる。さらに、金融機関における A T M（銀行自動預払機）、カーナビゲーションシステム、切符販売機等に搭載することも可能である。

【0 0 4 2】

また、再帰遷移ネットワーク作成部 2 において任意の言語（日本語、英語、ドイツ語等の各国言語）の文脈自由文法を選択して用いることにより、所望の言語に対応した有限状態変換器 2 2 を作成することができ、さらに、その有限状態変換器 2 2 を用いて所望の言語に対応した漸進的構文解析装置 2 1 を構成することができる。

【0 0 4 3】

【実施例】

(実験方法)

上述した本実施形態の有限状態変換器作成装置 1 によって有限状態変換器を実際に作成し、この有限状態変換器を用いて漸進的構文解析装置 21 を作成した。そして、漸進的構文解析装置 21 における漸進的構文解析の効果を検討するために、解析実験を行った。実験に用いた計算機のスペックは、CPU Pentium (登録商標) 4 2 GHz, メモリ 2 GB である。実験における学習データセット及びテストデータセットには、構文木付き ATR 音声言語データベース (日本語対話) を用いた。学習データ (文法規則の適用頻度に関する統計情報) として、言語データベースからランダムに 9, 081 文を抽出し、そこから、文法規則とそれらの適用確率を獲得した。このとき、文法規則は 698 種類、品詞は 337 種類、範疇は 153 種類であった。一方、テストデータとして 1, 874 文を用いた。テストデータ中の文の平均単語長は 9.4 単語であった。また、有限状態変換器の弧の数の閾値を 15, 000, 000 に設定した。この値に設定したのは、有限状態変換器の作成時において、メモリをほぼ限界まで使用したためである。このとき、解析時に使用するメモリの量は 600 MB 程度であった。

【0044】

(実験結果)

まず、本実施形態の有限状態変換器 1 を用いた漸進的構文解析装置 21 (実施例 1 とする) と、従来技術における漸進的チャート解析を用いた構文解析装置 (比較例 1 とする) とをそれぞれ用いて構文解析を行い、解析速度と精度とについて比較した。実施例 1 の有限状態変換器は、 $N=3$ としたときの文法規則の適用確率を使用して置き換え優先度を計算し、置き換える順序を決定した。ただし、 N は確率の計算に用いた文法規則の組が N 項組であることを示す。さらに、非終端記号をラベルとする弧を除去した。比較例 1 の漸進的チャート解析については、有限状態変換器作成に用いた文法規則の適用確率と同じ考えに基づき、ボトムアップ解析用に条件付き確率を求めて利用した。このとき、文法規則を適用するごとに、適用確率の積を計算し、その値が $1E-12$ を越えた場合には、それ以上の規則の適用を取りやめた。さらに、置き換える未決定項への到達可能性を用いて、文法規則の適用を制御した。さらに、実施例 1 の構文解析装置及び比較例

1の構文解析装置とも、1単語あたりの解析時間を10秒に制限し、その時間を越えた場合には、その単語についての解析を終了し、次の単語の解析へと進ませた。実施例1及び比較例1のそれぞれの構文解析装置における1単語あたりの解析時間、及び正解率を表2に示す。ただし、正解率は、文全体に対して得られた解析結果の中に、正解の構文木が存在した文の割合(%)である。正解の構文木は、文にあらかじめ付与されている構文木とした。

【表2】

実施例1と比較例1(漸進的チャート解析)との比較実験結果

	解析時間 (秒/word)	正解率 (%)
実施例1	0.05	87.5
比較例1 (漸進的チャート解析)	2.82	33.4

【0045】

実験結果より、実施例1の漸進的構文解析装置を用いることによって、比較例1よりも高速に解析できることがわかった。さらに、日本語の発話速度が1単語あたり0.25秒程度であるのに対し、実施例1の漸進的構文解析装置における解析速度は0.05秒となっており、発話速度を上回っている。これは、実施例1の漸進的構文解析装置が実時間での漸進的構文解析に有効であることを示している。

また、計算回数について比較するため、それぞれの解析方法について、1単語あたりの計算回数について調査した。有限状態変換器を用いた実施例1による解析については、状態を遷移して構文木を作成するときに1回の計算と数え、比較例1の漸進的チャート解析では、文法規則を適用するとき、及び、項を置き換えるとき、それぞれ1回の計算と数えた。その結果、1単語あたりの計算回数は、実施例1では1,209回、比較例1では、36,300回であり、実施例1では比較例1よりも計算回数が大幅に少なくなっていることから、有限状態変換

器を用いることによって構文解析処理を高速化できることがわかった。

【0046】

次に、有限状態変換器を使用した漸進的構文解析装置に関し、置き換え優先度を使用して作成した有限状態変換器を用いた実施例 2 及び 3 と、置き換え優先度を使用せずに作成した従来技術における有限状態変換器を用いた比較例 2 について、構文解析結果の正解率を比較する実験を行った。ここで、実施例 2 は、非終端記号をラベルに持つ弧の除去を実施しないで作成した有限状態変換器を用いた場合であり、実施例 3 は、弧の除去を実施して作成した有限状態変換器を用いた場合である。また、各実施例 2, 3 について、それぞれ文法規則の適用確率の条件の数を $N=0$ から $N=4$ まで変化させて有限状態変換器の作成を行った。実験結果を図 13 に示す。ただし、 N は、文法規則適用確率の規則条件数を表す。

実験結果から、有限状態変換器作成に置き換え優先度を利用した実施例 2, 3 の正解率は、利用しなかった比較例 2 に比べかなり向上しており、置き換え優先度を用いた弧の置き換え順序の制御は、有効であることがわかった。また、非終端記号の弧を除去した有限状態変換器を用いた実施例 3 は、弧の除去を行わなかった有限状態変換器を用いた実施例 2 よりも正解率が向上している。従って、いずれの実施例についても置き換え優先度を用いない比較例 2 よりも正解率が向上しており、さらに、置き換え優先度と非終端記号の弧の除去とを組み合わせることによって、80%後半の正解率を達成できることがわかった。また、文法規則の適用確率の条件数 N を 0 から 4 まで増加させるに従って正解率が向上していることがわかる。

【0047】

【発明の効果】

以上詳述したように、本発明の有限状態変換器作成装置、プログラム、記録媒体、作成方法によれば、文法規則の適用頻度に関する統計情報に基づく弧の置き換え優先度が高い弧から順に置き換え操作が適用されるので、限られた大きさで、より多くの文を解析可能な有限状態変換器を、確実に作成することができるという効果を奏する。

また、本発明の漸進的構文解析装置によれば、文法規則の適用頻度に関する統

計情報に基づく弧の置き換え優先度が高い弧から順に置き換え操作が適用された有限状態変換器を備え、接続処理手段が、その有限状態変換器へ単語を入力する度に状態遷移に伴って出力される構文木を順次接続するように構成されているので、文脈自由文法を近似変換した限られた大きさの有限状態変換器を用いて、より多くの文について漸進的に構文解析を行うことができるという効果を奏する。

【図面の簡単な説明】

【図 1】 本発明の一実施形態における有限状態変換器作成装置の全体構成を示すブロック図である。

【図 2】 文法規則の集合を表す P_X の一例を示す図である。

【図 3】 再帰遷移ネットワーク中の M_X の一例を示す図である。

【図 4】 再帰遷移ネットワークにおける状態の統合を説明する図である。

【図 5】 最初に与えられる初期有限状態変換器 M_0 を示す図である。

【図 6】 弧の置き換え操作の一例並びに弧と節点との対応関係を示す図である。

【図 7】 節点の導出における文法規則適用の過程を示す図である。

【図 8】 構文木から獲得される文法規則の組の一例を示す図である。

【図 9】 弧の連続的な除去方法を説明する図である。

【図 10】 本実施形態の漸進的構文解析装置の全体構成を示すブロック図である。

【図 11】 構文解析の一例を示す図である。

【図 12】 出力記号列が表す構文木の一例を示す図である。

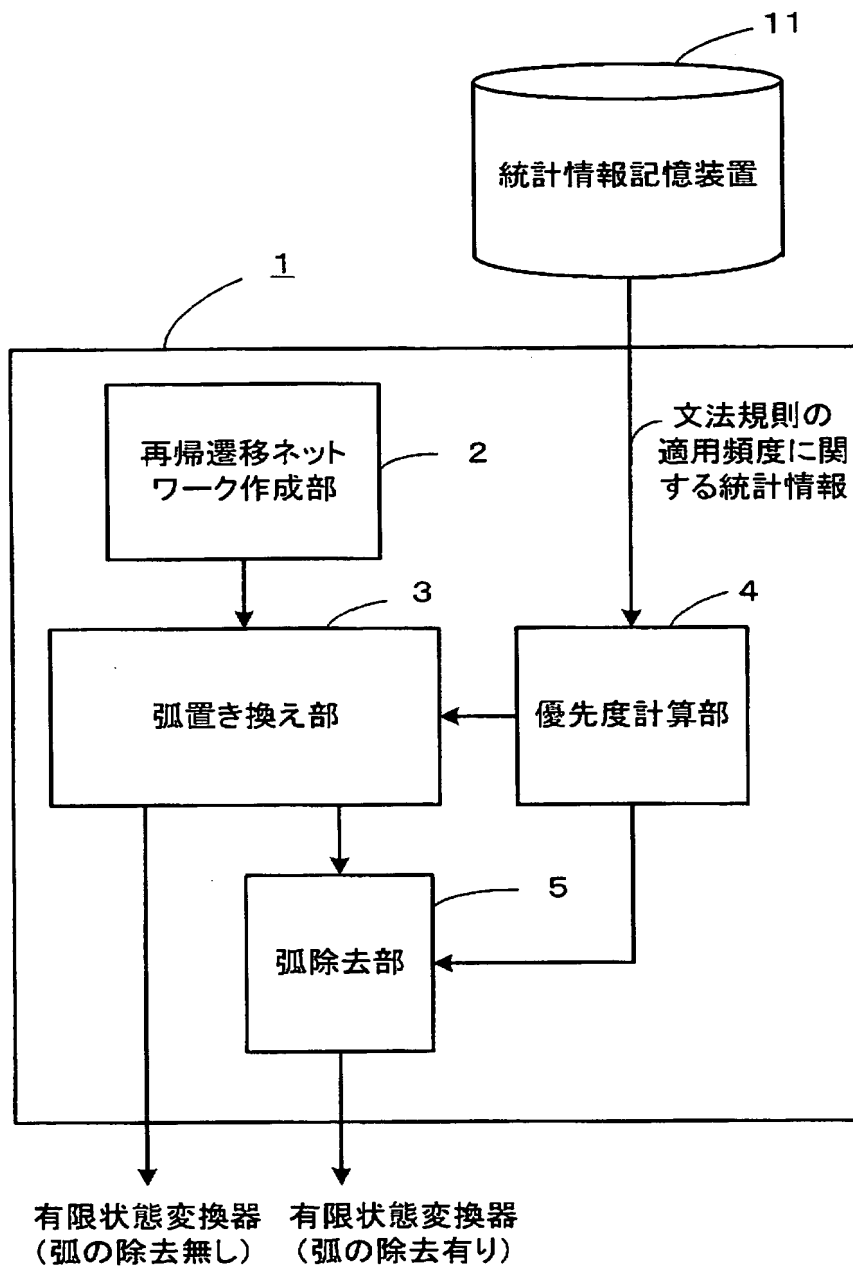
【図 13】 構文解析の実験結果（正解率）を示すグラフである。

【符号の説明】

1…有限状態変換器作成装置、2…再帰遷移ネットワーク作成部（再帰遷移ネットワーク作成手段）、3…弧置き換え部（弧置き換え手段）、4…優先度計算部（優先度計算手段）、5…弧除去部（弧除去手段）、21…漸進的構文解析装置、22…有限状態変換器、23…接続処理部（接続処理手段）。

【書類名】 図面

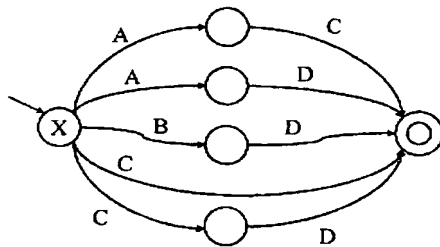
【図 1】



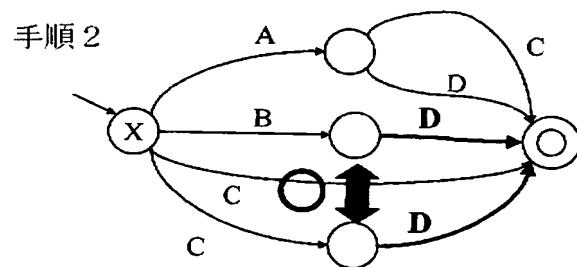
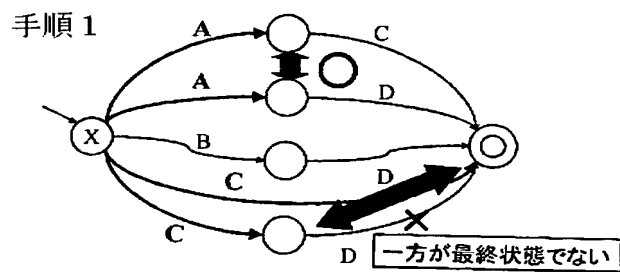
【図 2】

$X \rightarrow A C$
$X \rightarrow A D$
$X \rightarrow B D$
$X \rightarrow C$
$X \rightarrow C D$

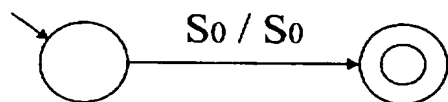
【図 3】



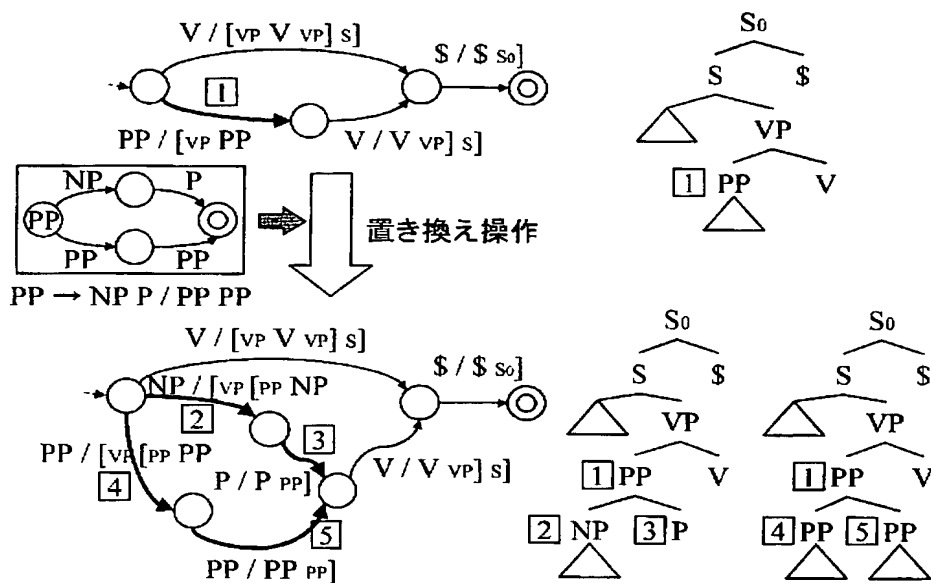
【図 4】



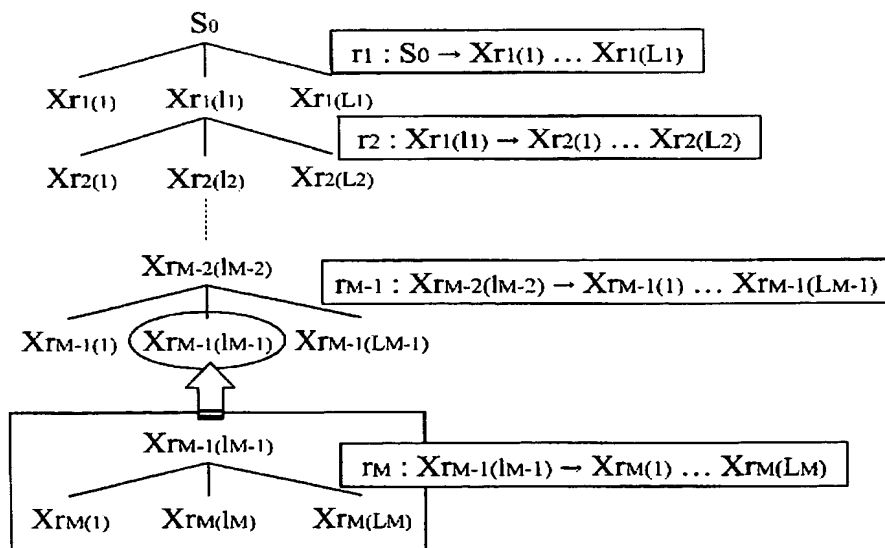
【図 5】



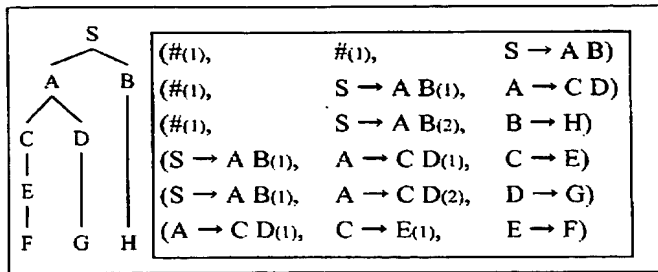
【図 6】



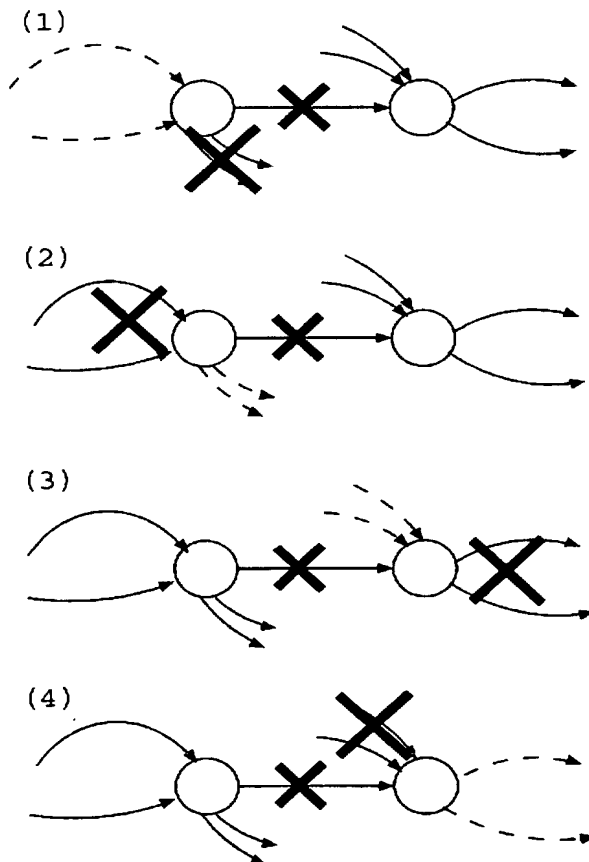
【図 7】



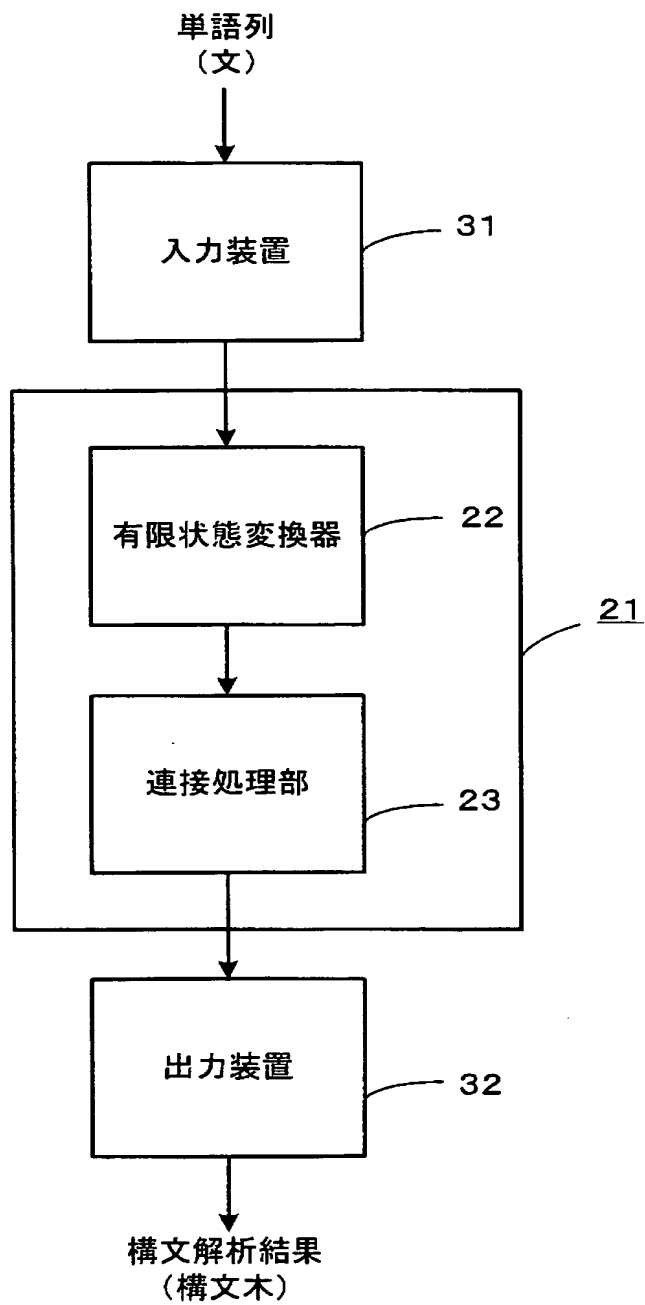
【図 8】



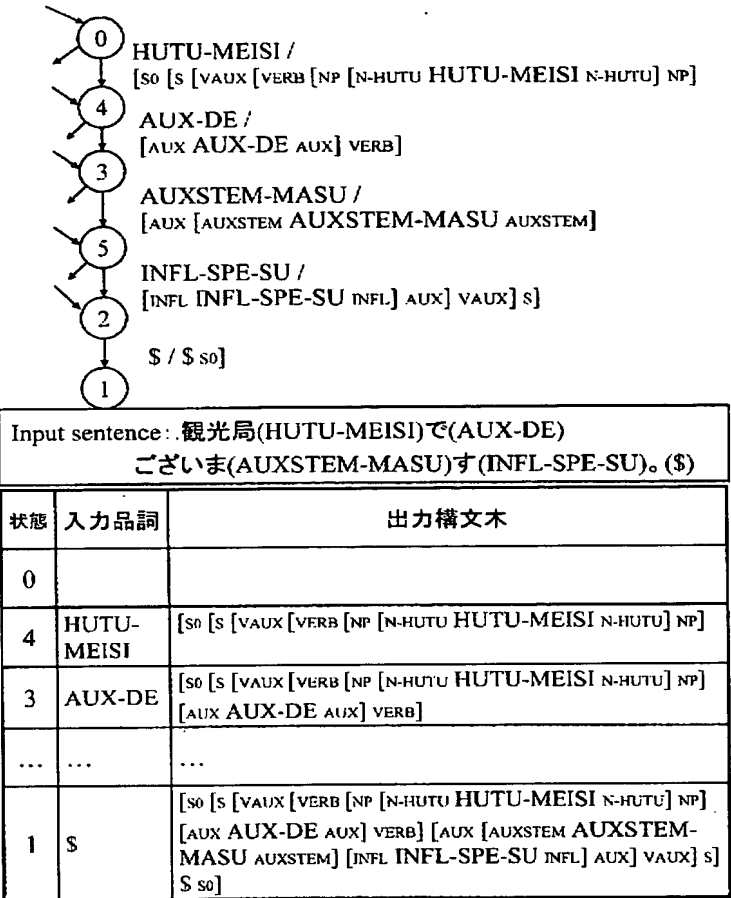
【図 9】



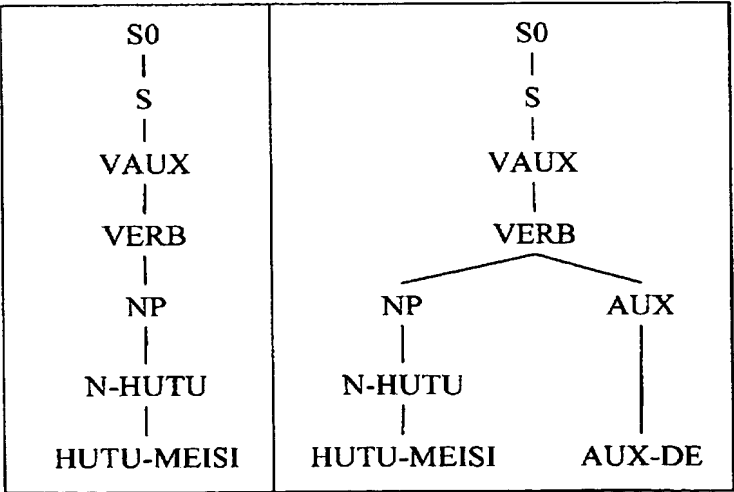
【図 10】



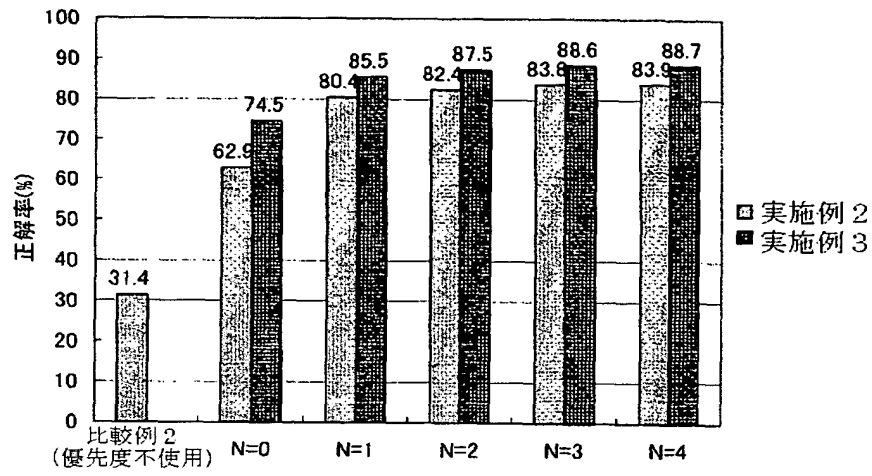
【図 1 1】



【図 1 2】



【図 13】



【書類名】 要約書

【要約】

【課題】 より多くの文について漸進的に構文解析可能な有限状態変換器の作成装置、プログラム、記録媒体、作成方法、及び漸進的構文解析装置を提供する。

【解決手段】 有限状態変換器作成装置 1 は、再帰遷移ネットワークを作成する再帰遷移ネットワーク作成部 2 と、有限状態変換器の弧をその入力ラベルに対応した再帰遷移ネットワーク中のネットワークで置き換える操作を再帰的に繰り返す弧置き換え部 3 と、文法規則の適用頻度に関する統計情報に基づいて弧の置き換え優先度を計算する優先度計算部 4 とを備え、弧置き換え部 3 は、弧の置き換え優先度が高い弧から順に置き換え操作を適用するので、限られた大きさで、より多くの文を構文解析可能な有限状態変換器を、確実に作成することができる。

【選択図】 図 1

特願 2003-060681

出願人履歴情報

識別番号

[598091860]

1. 変更年月日

1998年 7月 9日

[変更理由]

新規登録

住 所

愛知県名古屋市中区栄二丁目10番19号

氏 名

財団法人名古屋産業科学研究所